VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

STATISTICAL ANALYSIS OF FEATURES AND CLASSIFIERS IN IDENTIFYING NODULES AND ITS T STAGING IN LUNG CT IMAGES

G. Niranjana[1] and M. Ponnavaikko[2]
[1]Department of Computer Science Engineering, SRM University, Chennai, India
[2]Bharath University, Chennai India
E-Mail: niran_janag@yahoo.com

ABSTRACT

Lung cancer is the most common disease with greater morality rate. Computed Tomography (CT) images are used for early diagnosis of lung cancer with the help of CAD system. Selection of effective feature set and proper classifier for medical images where machine learning techniques are used is a challenging task. Texture analysis of computed tomography (CT) images is one of the important preliminary stages in the detection and classification for lung cancer. The image texture is characterized by Haralick texture with variety of statistical measures. The extracted texture feature values are used by a CAD to differentiate its type as benign or malignant. This paper aims to compare experimental results of 18 features extracted by using Gray Level Co occurrence Matrix (GLCM) and analyses the different classifiers that can be used for classification of nodule as benign or malignant. Measuring the statistical parameters of the nodule is crucial for determining the T stage of the nodule. This paper also analysis the statistical parameters and reported the contribution of minimal feature set for classification and staging. GLCM features are used for classification and Geometric features used are used for T staging. For these analysis 23 images dataset of different types of cancer is used.

Keywords: lung cancer, GLCM, feature set, classifier, nodule type, T staging.

1. INTRODUCTION

1.1 General

Lung cancer is a disease that occurs because of unwanted growth of tissues of the lungs. Among all the cancers, the lung cancer causes the maximum cases of deaths in Men and Women [1]. Early diagnose of the lung tumor can increase the survival rate of 1 to 5 years. In a normal case, the human body checks and maintains the growth of cells in order to produce new cells whenever they are required. The unbalance of the system results in uncontrolled division and proliferation of cells due to which a mass is formed, known as a tumour. Tumour which grows aggressively and spread into the other parts of the body is known as malignant. Various computer-aided diagnosis (CAD) systems have been designed for the early diagnose of lung tumor [1].

In medical Imaging different types of images are being used, but for the detection of lung cancer Computed Tomography (CT) images are being preferred because of better clarity, low noise and less distortion. The CAD system for processing lung CT images consists of several stages.

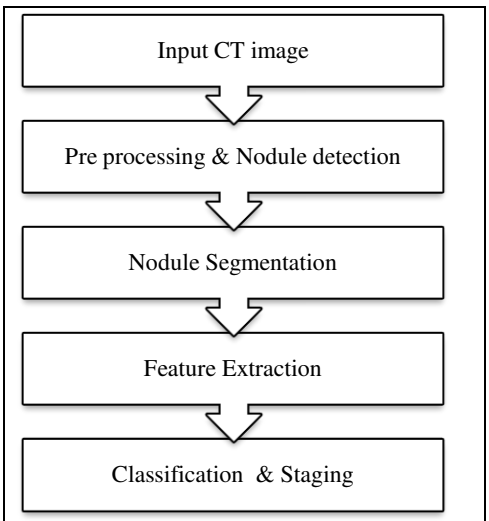

Figure-1. Stages of Lung Cancer.

Among these, Pre processing or lung segmentation, Nodule detection, Nodule segmentation, Feature Extraction and Classification are mandatory stages. Typical CAD system for lung cancer is shown in Figure-1. The input to the CAD system is the lung CT image. A lung segmentation step is used as a pre processing step to reduce the search space for lung nodules. Nodule detection step identifies the locations of lung nodules. The detected nodules are segmented by using any of the available segmentation algorithms [23]. Then, the desired features are extracted from the nodules.

These features are used to classify the nodule as benign or malignant and to stage it [24].

The classification of detected nodule into benign or malignant is based on the texture features [11]. Qian Zhao, Chang-Zheng Shi, Liang-Ping Lu in their paper proposed the role of texture features in the diagnosis of lung nodules [5]. In their study, the role of textural features in the differentiation of nodules had been investigated. Segmentation of pulmonary nodules and extraction of their spatial structures and gray-scale for improving the specificity were quantitatively described.

Texture features of segmented nodules were analyzed by the method of grey level co-occurrence matrix (GLCM) by Qian Zhao *et al* [25]. The variables of texture features include contrast, homogeneity, correlation, energy and entropy.

The proposed method uses DICOM images of lung CT images as input images. Contrast enhancement and histogram equalization techniques are used for image enhancement. Further the region of interest i.e. the lung nodule in the CT image is segmented using Random walker algorithm [6]. The size of the nodule, its shape, volume of the detected tumors and Gray Level Co occurrence Matrix are used for extraction of features. Based on these features Naive Bayes Classifier is used for classification of the tumours into Benign and Malignant tumours and Support Vector Machine (SVM) for corresponding T stage and the results are compared with other existing classifiers to show accuracy.

The existing techniques are used only to classify the tumours as benign or malignant tumours. The present research aims to classify the tumours detected into cancer stages based on T staging according to TNM staging classification [8] and verified using performance measures. It will help radiologist to improve the diagnosis efficiency by calculating the quantity of nodule growth in each stage accurately [7].

1.2 Feature extraction techniques

Image feature extraction stage is an important stage that uses algorithms and techniques to detect various desired portions or shapes. The selected features of the affected part must be extracted [16]. Grey-Level Co-occurrence Matrix (GLCM) parameters are used for analysing the texture of the image [25]. GLCM is a statistical method proposed by Haralick [3], and it contains many statistical features. GLCM parameters are obtained by calculating the occurrence of pairs of specific values with specific relationship in an image. These features are generated by calculating the features for each one of the cooccurrence matrices obtained by using the directions 0°, 45°, 90°, and 135°, then averaging these four values [5].They are used to characterise the texture of the image. In this work we used 18 features and they are autocorrelation, contrast, correlation, Cluster_promin, cluster_shade, dissimilarity, energy, entropy, homogeneity, Max_prob, sumOfSq, sumAvg, sumVar, sumEnt, diffVariance, diffEntropy, infmeas_corr, invdiffN and invdiffMN [10, 14].

Geometric parameters such as Area, Major and Minor Axis, Perimeter, Eccentricity and Diameter are used for T stage identification. This paper concentrates on feature extraction and analysis of features of 23 images of different types of cancer that may be used for classifying the nodule as benign or malignant and the corresponding T staging.

The features that are extracted and used for classification are listed in the Appendix [15].

1.3 Classification techniques

1.3.1 Support Vector Machine (SVM)

Hiram Madero Orozco *et al* [13] stated that Support Vector Machine (SVM) is a commonly used classification method for medical images [21]. SVM was proposed by Vapnik [19]. Support vector machines are supervised learning models that are associated with learning algorithms for analyzing data used for classification. Hiram Madero Orozco *et al* [13] proposed an efficient methodology for lung nodule classification without the stage of segmentation. Eight texture features were extracted from the histogram and the gray level co-occurrence matrix (with four different angles) after image acquisition for each CT image. Support vector machine (SVM), used to classify lung tissues into two classes: with lung nodules and without lung nodules. They obtained the better reliability results with 90◦ and 135◦ of the GLCM in their work.

The basic SVM is a non-probabilistic binary classifier that takes a set of input data and for each given input, predicts which of two classes forms the input [18]. From the given set of training examples, each belonging to one of the output categories, an SVM training algorithm builds a model that categorizes the given input to one of the output classes.

1.3.2 Naïve Bayes classifier

Bayesian classifier (Naive Bayes Classifier) is a probabilistic classification method based on threshold parameters rather than based on trial and error testing. Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the assumption of independence between every pair of features. This classifier classifies the objects on the basis of probabilities of detected features. The probabilities of the class membership are assigned on the basis of Bayes theorem [20]. From a Bayesian viewpoint, a classification problem can be written as the problem of finding the class with maximum probability given a set of observed attribute values. The probability is calculated using the Bayes theorem. The extracted features need to be classified as one of the stages which will be useful to the physician to plan the treatment. A naive Bayes classifier assumes that the value of a particular feature is unrelated

to the presence or absence of any other feature. The advantage of Naive Bayes classifiers is that they can be extremely fast compared to other more sophisticated methods. Bayesian classifier is used to classify the detected tumours into appropriate stage.

1.3.3 Sequential Minimization Optimization (SMO)

Platt's sequential minimization optimization (SMO) algorithm is a fast iterative algorithm [25]. This algorithm replaces all missing values and transforms nominal attributes into binary ones. It normalizes all the input attributes to the classifier by default. The coefficients in the output are also based on the normalized data and not on the original data. Multi-class problems are solved using pairwise classification of SMO.

1.3.4 T staging of the nodule

The three algorithms discussed above are tested and the results are compared separately for lung nodule type detection and T-staging of the nodule. T staging is identified using the parametric values given in Table-1.

Table-1. Criteria for T descriptor.

Stage parameter	T1	T2	T3	T4
Equiv diameter (cm)	5 to 7	7 to 11	11 to 21	21 to 42
Perimeter (pixels)	18 to 25	25 to 38	38 to 80	80 to 177
Area (pixels)	0 to 57	57 to 100	100 to 352	1446

2. PROPOSED APPROACH

2.1 Performance parameters

Performance measures are the yardsticks for evaluating any system. The comparison of SVM, SMO and Naive Bayes classifier classification algorithms is done on the basis of the following performance parameters:

a) Confusion matrix

A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix [22]. The following Table-2 shows the confusion matrix for a two class classifier:

Table-2. Confusion matrix predicted label

		Positive	Negative
Known label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Negative (FN)	True Negative (TN)

b) TP Rate

The True Positive (TP) rate is the proportion of samples that were classified as class x, which were originally belong to class x, i.e. benign nodule is identified as benign and malignant nodule is identified as malignant. It is equivalent to Recall. In the confusion matrix, this is the diagonal element divided by the sum over the relevant row. TP can be calculated with the following formula:

$$TP = TP/(TP + FN) \quad (1)$$

c) FP Rate

The False Positive (FP) rate is the proportion of samples which were classified as class x, but they belong to a different class. i.e., benign nodule is identified as malignant. In the confusion matrix, this is the column sum of class x minus the diagonal element, divided by the rows sums of all other classes. FP can be calculated using (2).

$$FP = FP / (FP + TN) \quad (2)$$

d) Correctly classified instances

Correctly classified instances are used in order to find out which algorithm correctly classifies maximum number of sample instances.

e) Incorrectly classified instances

Incorrectly classified instances are used in order to find out which algorithm incorrectly classifies maximum number of sample instances.

f) Root mean square error

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure. It is calculated as the differences between values predicted by a model or an estimator and the values actually observed. It is used to represent the degree of randomness of the pixels in the image.

g) Mean absolute error

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables.

Expressed in words, the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average.

h) Precision

Precision is the proportion of the predicted positive cases that were correct, as calculated using the equation given below:

$$\text{Precision } p = TP / (TP+FP) \quad (3)$$

i) Recall

Recall (true positive rate) is the proportion of positive cases that were correctly identified, as calculated using the equation below:

$$\text{Recall } r = TP / (TP+FN) \quad (4)$$

j) F-measure

F measure is a measure of a test's accuracy. It considers both the precision *p* and the recall *r* of the test to compute the score: *p* is the number of correct positive results divided by the number of all positive results, and *r* is the number of correct positive results divided by the number of positive results that should have been returned. The F_1 score can be interpreted as a weighted average of the precision and recall, where an F_1 score reaches its best value at 1 and worst score at 0.

2.2 Classification results for nodule type

a. Confusion matrix

Table-3 gives the confusion matrix for different types of classification of the nodule. The classes taken are Benign (B) and malignant (M).

Table-3. Confusion matrix for nodule type classification.

	Predicted class		
		B	M
Actual class	B	20	1
	M	2	1

(a) Using SVM

	Predicted class		
		B	M
Actual class	B	18	2
	M	2	2

(b) b) Using NaiveBayes

	Predicted class		
		B	M
Actual class	B	12	9
	M	1	2

(c) (c) Using SMO

b. Results of classification

Table-4 shows the summary of the classification and its results for performance measures for all the 3 classifiers with 24 sample images and 7 attributes. Table-4 gives the number of correctly classified instances. As could be seen this number of instances classified correctly is higher for SVM than SMO and Bayesian classifier.

Table-4. Summary of classification.

MEASURES	SMO Classifier	SVM classifier	NaiveBayes classifier
Correctly Classified Instances	14	21	20
Incorrectly Classified Instances	10	3	4
Mean Absolute Error	0.416	0.0652	0.3007
Root Mean Squared Error	0.311	0.284	0.395
Relative Absolute Error	69.733	17.54	80.873
Root Relative Squared Error	89.4	58.76	88.3244

Total Number of Instances 24
Total Number of Attributes 7

c. Detailed accuracy by class

The Table-5 below gives the detailed accuracy measures for classification. It is observed from the following table that the TP rate is high for SVM.

Table-5. Detailed accuracy by class for classification.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
SMO	0.683	0.345	0.83	0.583	0.653	0.754
Naïve Bayes	0.833	0.595	0.833	0.833	0.833	0.861
SVM	0.875	0.018	0.936	0.875	0.891	0.967

d. Comparison with other classifier algorithms

To find the performances of various algorithms, Graphs have been plotted with the minimal attribute set. It has been observed from the plotted vales that SVM correctly classified 96.7% of instances whereas SMO classifies 75.4%, and naive bayesclassifies 86.1% of instances. SVM classifier gives maximum number of correctly classified instances and SMO gives minimum number of correctly classified instances. Figure-2 shows the comparison.

Table-6. Sensitivity and specificity values for Naive Bayes, SVM and SMO.

	Naive Bayes classifier	SVM	SMO
Correct	0.861	0.967	0.754
Incorrect	0.139	0.033	0.246
RMSE	0.311	0.284	0.395

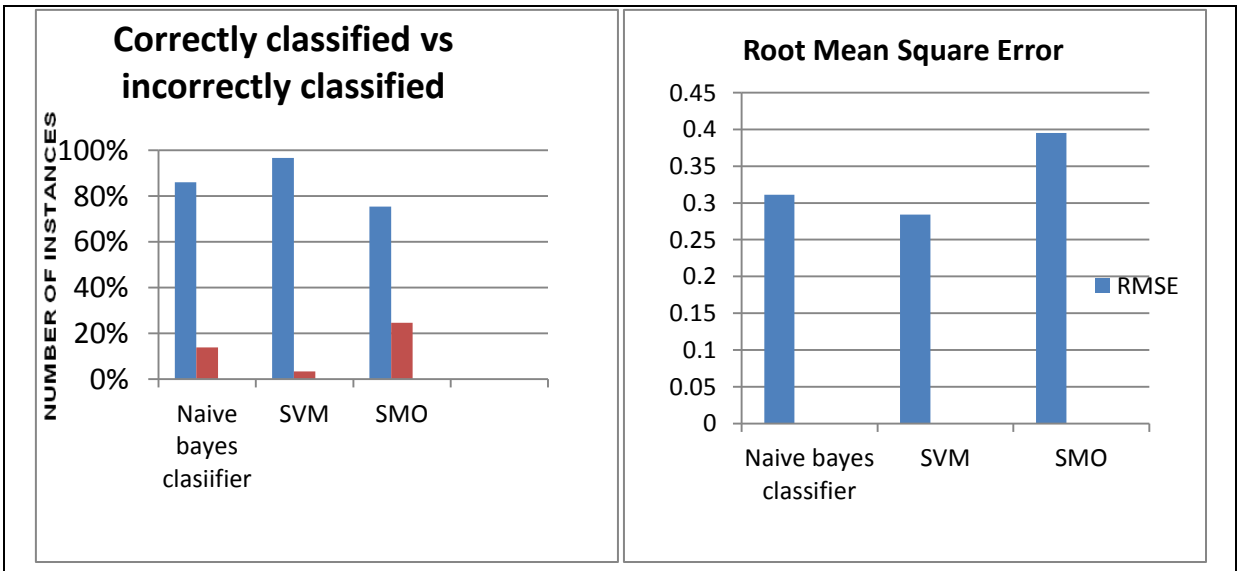

Figure-2. (a) Classification Algorithms comparison. (b) Root Mean Square Error comparison.

It has been observed from the plotted RMSE values that SVM classifier gives minimum value and other Classifiers give maximum value of Root Mean Square Error value.

2.3 classification results for T Staging

a. Confusion matrix

Table-7 gives the confusion matrix for different stage classification of the nodule.

Different T stages of the nodule are T1, T2, T3 or T4.

ARPN Journal of Engineering and Applied Sciences

Table-7. Confusion matrix for T-staging.

	Predicted class															
		T1	T2	T3	T4		T1	T2	T3	T4		T1	T2	T3	T4
Actual class	T1	5	1	0	0	T1	5	1	0	0	T1	4	1	1	0
	T2	1	5	0	0	T2	1	4	1	0	T2	1	4	1	0
	T3	0	0	8	0	T3	0	0	8	0	T3	1	0	7	0
	T4	0	0	0	3	T4	0	0	0	3	T4	0	0	0	3
	Naïve Bayes					(b) SVM					(c) SMO				

b. Results of T staging

Table-8 shows the summary of the classification and its results for performance measures for all the 3 classifiers with 23 sample images and 18 attributes. From the Table below it is observed that the number of correctly classified instances is higher for Naïve Bayes than SMO and SVM classifier.

Table-8. Summary of classification for T staging.

Measures	SMO classifier	SVM classifier	NaiveBayes classifier
Correctly Classified Instances	21	20	12
Incorrectly Classified Instances	2	3	11
Mean Absolute Error	0.0436	0.0652	0.3007
Root Mean Squared Error	0.208	0.2554	0.3839
Relative Absolute Error	11.7336%	17.54	80.873
Root Relative Squared Error	47.859	58.76	88.32

Total Number of Instances 23
Total Number of Attributes 7

c. Detailed accuracy By class

The Table-9 below gives the detailed accuracy measures for classification. It is observed from the following table that the TP rate is high for Bayesian classifier.

Table-9. Detailed accuracy by class for T staging by Naive Bayes classifier.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
SVM	0.833	0	1	0.833	0.909	0.902
NaiveBayes	0.87	0.054	0.866	0.87	0.865	0.972
SMO	. 0.522	0.184	0.497	0.522	0.508	0.743

d. Comparison with other classifier algorithms

To find the performances of various algorithms, Graphs have been plotted with the minimal attribute set. It has been observed from the plotted values that Bayesian classifier correctly classified 92.3% of instances whereas SMO classifies 73.9%, and SVM classifies 60.9% of instances. SVM classifier gives maximum number of correctly classified instances and SMO gives minimum number of correctly classified instances. Figure-2 shows the comparison.

Table-10. Sensitivity and specificity values for Naive Bayes, SVM and SMO.

	Naive Bayes classifier	SVM	SMO
Correct	0.923	0.609	0.739
Incorrect	.077	0.391	0.261
RMSE	0.208	0.442	0.355

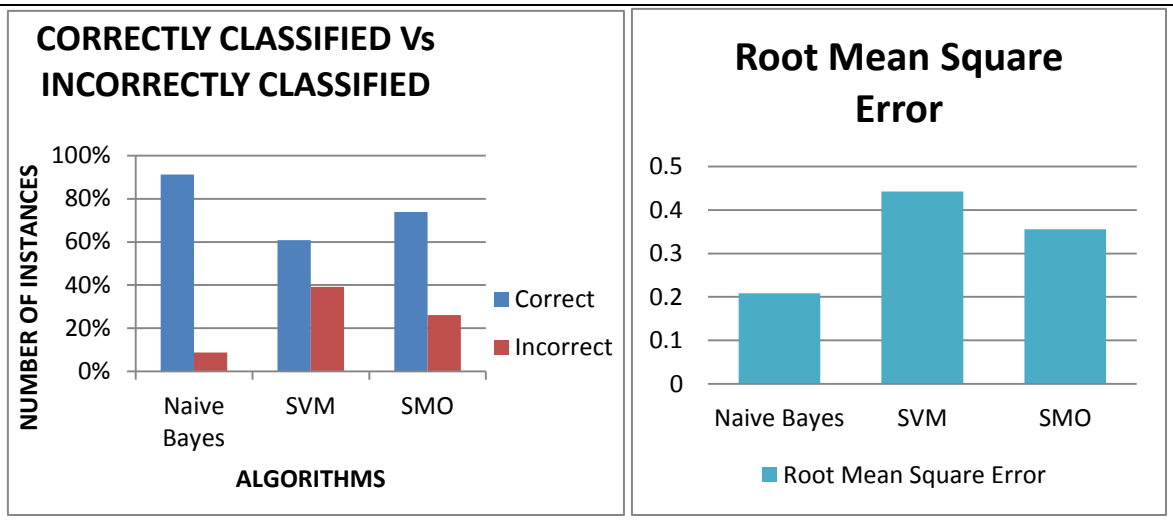

Figure-3. (a) Classification algorithms comparison. (b) Root Mean Square Error comparison.

It has been observed from the plotted vales that Naive Bayes correctly classified 92.3% of instances whereas SMO classifies 60.86%; SVM classifies 73.91% of instances. Naive Bayes classifier gives maximum number of correctly classified instances and SMO gives minimum number of correctly classified instances. It has been observed from the plotted values that Naive Bayes classifier gives minimum value and other SVM Classifier gives maximum value of Root Mean Square Error.

3. PERFORMANCE ANALYSIS

a. Minimal feature set

Information gain algorithm is applied and attributes are arranged in the order of significance. Equivalent Diameter attribute shows the highest priority. The classification algorithms are applied taking one attribute at a time in descending order of priority i.e. Equivalent diameter, Major Axis length, Area, Perimeter, Minor Axis Length, and Eccentricity. Values of different parameters have been compared in order to evaluate which algorithm gives better results in case of Lung CT-Scan Images.

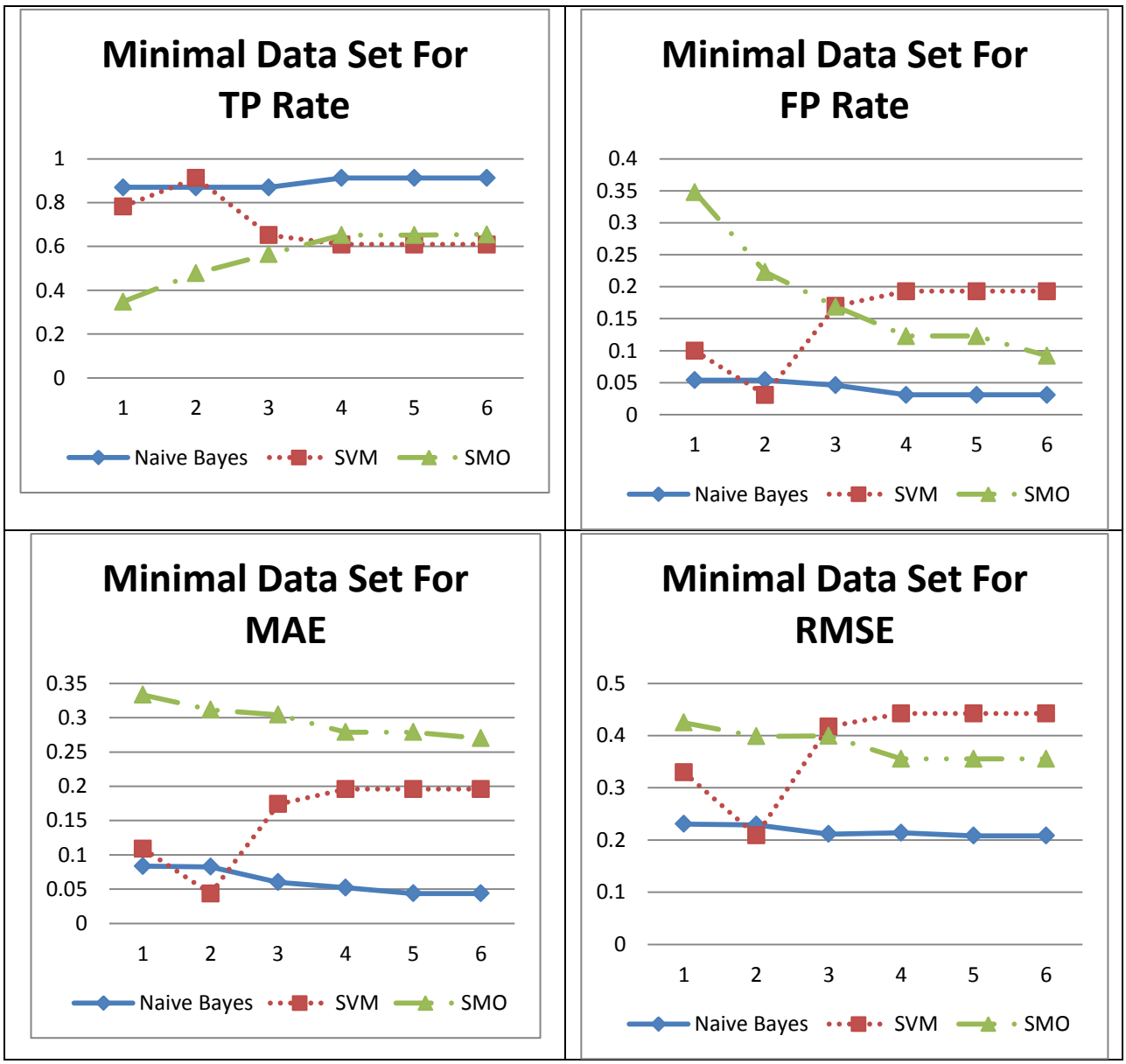

Figure-4. (a) Minimal feature set using TP Rate. (b) Minimal feature set using FP Rate. (c) Minimal feature set using Mean Absolute Error (d) Minimal feature set using Root Mean Square Error.

From the graphs it is found that the attributes equivalent diameter, major Axis length, Area, Perimeter form the Minimal Attribute Set. This minimal attribute set gives optimal results for the classifying algorithms being used.

CONCLUSIONS

The statistical analysis of features and classifiers for identifying nodules and its T staging in lung CT images performed with the three classifiers SVM, Bayes and SMO were presented in this paper. The following results were observed in this study: The entropy, contrast and correlation of malignant nodules were higher than those of benign nodules. However, the energy, homogeneity and correlation were higher in benign nodules than in malignant nodules. Therefore, there were statistically significant differences in the variables of texture features. Thus, the heterogeneity or complexity of malignant nodules was higher than those of benign nodules, while more homogenous and uniform appearances were observed for benign nodules. With respect to classification of nodule type Support Vector Machine gives better results than Bayesian and SMO classifier. For T staging classification prominent results are obtained using Naïve Bayes classifier than SVM and SMO classifiers.

REFERENCES

[1] JaspinderKaur, NidhiGarg, Daljeet Kaur. 2014. A survey of Lung Cancer Detection Techniques on CT scan Images. International Journal of Scientific and Engineering Research. ISSN: 2229-5518, 5(6): 377-380.

[2] Fritz Albregtsen. Statistical Texture Measures Computed from Gray Level Coocurrence Matrices.

[3] L. Soh and C. Tsatsoulis. 1999. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. IEEE Transactions on Geoscience and Remote Sensing. 37(2): 80-795.

[4] F. I. Alam, R. U. Faruqui. 2011. Optimized Calculations of Haralick Texture Features. European Journal of Scientific Research. 50(4): 543-553.

[5] D A. Clausi. 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Remote Sensing. 28(1): 45-62.

[6] Sundaresh Ram and Jeffrey J. Rodríguez. 2013. Random Walker Watersheds: A New Image Segmentation Approach. in Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Vancouver, Canada. pp. 1473-77.

[7] Qian Zhao, Chang-Zheng Shi, Liang-Ping Luo. 2014. Role of the texture features of images in the diagnosis of solitary pulmonary nodules in different sizes. Chinese Journal of Cancer Research. ISSN: 1000-9604, pp. 451-458.

[8] S. Tsim, C.A. O'Dowd, R. Milroy, S. Davidson. 2010. Staging of non-small cell lung cancer (NSCLC): A review. Elsevier, Respiratory Medicine. 104(12): 1767-1774.

[9] Mehrdad J. Gangeh, LaugeSørensen, Saher B. Shaker, Mohamed S. Kamel, Marleen de Bruijne and Marco Loog. 2010. A Texton-Based Approach for the Classification of LungParenchyma in CT Images. Springer LNCS 6363, pp. 595-602.

[10] Xu Y., Sonka M., McLennan G., Guo J., Hoffman E.A. 2006. MDCT-based 3-D Texture Classification of Emphysema and Early Smoking Related Lung Pathologies. IEEE Trans. Med. Imag. 25(4): 464-475.

[11] Uppaluri R., Mitsa T., Sonka M., Hoffman E.A., McLennan G. 1997. Quantification of Pulmonary Emphysema from Lung Computed Tomography Images. Amer. J. Respir. Crit. Care Med. 156(1): 248-254.

[12] Ms. Swati P. Tidke, Prof. Vrishali A. Chakkarwar. 2012. Classification of Lung Tumor Using SVM. International Journal of Computational Engineering Research (ijceronline.com). 2(5).

[13] Hiram Madero Orozco, Osslan Osiris Vergara Villegas. 2013. Lung Nodule Classification in CT Thorax Images using Support Vector Machines. 12th Mexican International Conference on Artificial Intelligence. pp. 277-283.

[14] Kesav Kancherla, Srinivas Mukkamala. 2013. Early Lung Cancer Detection using Nucleus Segmentation based Features. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 91-95.

[15] Prashant Naresh, Dr. Rajashree Shettar. 2014. Image Processing and Classification Techniques for Early Detection of Lung Cancer for Preventive Health Care: A Survey. Int. J. of Recent Trends in Engineering and Technology. Vol. 11.

[16] Hang See Pheng; Siti M. Shamsuddin. 2013. Texture classification of lung computed tomography images. Proc. SPIE 8768, International Conference on Graphic and Image Processing (ICGIP 2012), 87683Z.

[17] Hiram Madero Orozco, Humberto de Jes´us Ochoa Dom´ınguez. 2013. Lung Nodule Classification in CT Thorax Images using Support Vector Machines. 12th Mexican International Conference on Artificial Intelligence.

[18] Mylene T., Bradley S. and Jane P. 2010. Multidetector CT of solitary pulmonary nodules. International Journal of Engineeringand Technology. 20(1): 9-23.

[19] Cortes C, Vapnik V. 1995. Support-vector networks [J]. Machine Learning. 20: 273-297.

[20] Duda Hart and Stork, Pattern Classification, 2nd Edition, Wiley India, 2010.

[21] Jhilam Mukherjee, Amlan Chakrabarti, Soharab Hossain Skaikh. 2014. Automatic Detection and

Classification of Solitary Pulmonary Nodules from Lung CT Images. Fourth International Conference of Emerging Applications of Information Technology.

[22] Nadav David Marom, Lior Rokach, Armin Shmilovici. 2010. Using the Confusion Matrix for Improving Ensemble Classifiers. IEEE 26-th Convention of Electrical and Electronics Engineers in Israel.

[23] Ayman El-Baz, Garth M. Beache, Georgy Gimel'farb, Kenji Suzuki, Kazunori Okada, Ahmed Elnakib, Ahmed Soliman, Behnoush Abdollahi. 2013. Computer-Aided Diagnosis Systems for Lung Cancer: Challenges and Methodologies. Hindawi Publishing Corporation International Journal of Biomedical Imaging, Vol. 2013.

[24] Qian Zhao, Chang-Zheng Shi, Liang-Ping Luo. 2014. Role of the texture features of images in the diagnosis of solitary pulmonary nodules in different sizes. Chinese Journal of Cancer Research, ISSN: 1000-9604, pp. 451-458.

[25] John C. Platt. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft Research, Technical Report MSR-TR-98-14.

Appendix A

The following features are used:

- **Area (A)**

It shows the actual number of pixels in the ROI.

$$A=\sum_{i,j} f(i,j) \tag{5}$$

For a binary image, if 1 represents object and 0 represents background, then the area A , is the number of f(i,j) = 1.

- **Major and minor axis of ROI**

The equivalent elliptical major and minor axis of irregular region is defined as:

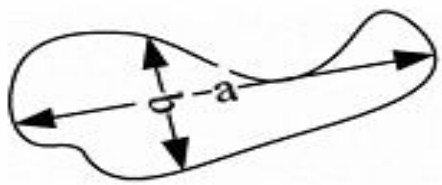

Figure-5. Major axis and Minor axis of a region.

Major axis,

$$a=2\times\left[\frac{2\left(\mu_{20}+\mu_{02}+\sqrt{(\mu_{20}+\mu_{02})^2+4\mu_{11}^2}\right)}{\mu_{00}}\right]^{1/2} \tag{6}$$

Minor axis,

$$b=2\times\left[\frac{2\left(\mu_{20}+\mu_{02}-\sqrt{(\mu_{20}+\mu_{02})^2+4\mu_{11}^2}\right)}{\mu_{00}}\right]^{1/2} \tag{7}$$

where μ is the central moment. The major axis 'a' is seen as the diameter of ROI .

- **Perimeter**

Perimeter property is calculating the distance between each adjoining pair of pixels around the border of the region. If the image contains discontinuous regions, then it returns unexpected results. Perimeter is calculated by counting the pixels contained only in the boundary.

- **Eccentricity**

The eccentricity is defined as the ratio of the distance between the focus of the ellipse and its major axis length.

$$\text{Eccentricity}=\text{Minor Axis Length}/\text{Major Axis Length} \tag{8}$$

- **Equivalent diameter**

It is defined as the diameter of a circle with the same area as the ROI.

$$\text{Equivalent Diameter}=\frac{\sqrt{4*\text{Area}}}{\sqrt{\pi}} \tag{9}$$

- **Energy:** It is the summation of squared elements in the GLCM and its value ranges between 0 and 1.

$$\text{Energy}=\sum_{k=0}^{n} p^2(i,j) \tag{10}$$

- **Contrast:** It is the measure of contrast between an intensity of pixel and its neighboring pixels over the whole ROI. where N is the number of different gray levels.

$$\text{Contrast} = \sum_{i=0}^{n} \sum_{j=0}^{n} (i-j)^2 p(i,j) \quad (11)$$

- **Homogeneity:** It is the measure of closeness of the distribution of elements in the GLCM to the GLCM of each ROI and its Value ranges between 0 and 1.

$$\text{Homogeneity} = \sum_{i,j}^{n} \frac{p(i,j)}{1+|i-j|} \quad (12)$$

- **Correlation:** It is the measure correlation of pixel to its neighbor over the ROI.

$$\text{Correlation} = \sum_{j=1}^{n} \sum_{i}^{n} \frac{p(i,j)-\mu r \mu c}{\sigma r \sigma c} \quad (13)$$